**Descriptive Statistics**

**Open this link and download the file(s)**

**https://drive.google.com/open?id=1PEubyOUwogSXWxgEhzp708mJmwBrQmIS**

**Population (N) vs Sample (n)**

-Sample should have randomness and representativeness

**Types of Data:**

-Categorical (Yes/No, Car Brands)

-Numerical

-Discrete & Continuous

**Levels of measurement:**

-Qualitative

-Nominal -> cannot be represented in Order (Car Brands, Four Seasons)

-Ordinal -> can be put in order (Rating a meal)

-Quantitative

-Ratio -> has a true 0 (number of bottles, Length, degree K)

-Interval -> doesn’t have a true 0 (Fahrenheit, Celcius)

**Visualization Techniques:**

-Categorical Data:

-Frequency Distribution Tables

-Bar Charts

-Pie Charts

-Pareto Diagrams

-Numerical Data:

-Frequency Distribution Tables

-Histograms

**Measures of Central Tendency:**

-Mean (average) (µ for population or for sample) -> sum of all elements/N

-Keep Outliers in mind (Can affect result)

-Median -> element at position n+1/2

-Mode -> Most number of occurrences

**Measures of Asymmetry:**

-Skewness -> whether data is concentrated on one side

-mean>median -> positive Skew

-mean=median=mode -> zero skew

-mean<median -> negative skew

**Measures of Variability:**

-Variance -> Dispersion of a set of data points around their mean

-Population Variance:

-Sample Variance:

-Standard Deviation (Like Variance but under root in order to deal with large values of variance)

- Population SD:

-Sample SD:

-Coefficient of Variation (aka relative standard deviation)

-SD/mean

**Measures of Relationship between variables:**

-Covariance

-Population covariance:

-Sample covariance:

-Linear correlation coefficient ->

-Correlation doesn’t imply causality

-Correlation of two variables is symmetrical i.e x,y=y,x